Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms

Boubacar Mainassara, Yacouba

7 December 2009

Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms

Y. Boubacar Mainassara[a]

[a] *Université Lille III, EQUIPPE, BP 60 149, 59653 Villeneuve d'Ascq cedex, France.*

Abstract

We consider portmanteau tests for testing the adequacy of structural vector autoregressive moving-average (VARMA) models under the assumption that the errors are uncorrelated but not necessarily independent. The structural forms are mainly used in econometrics to introduce instantaneous relationships between economic variables. We first study the joint distribution of the quasi-maximum likelihood estimator (QMLE) and the noise empirical autocovariances. We then derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. We deduce the asymptotic distribution of the Ljung-Box (or Box-Pierce) portmanteau statistics in this framework. It is shown that the asymptotic distribution of the portmanteau tests is that of a weighted sum of independent chi-squared random variables, which can be quite different from the usual chi-squared approximation used under independent and identically distributed (iid) assumptions on the noise. Hence we propose a method to adjust the critical values of the portmanteau tests. Monte carlo experiments illustrate the finite sample performance of the modified portmanteau test.

Key words: Goodness-of-fit test, QMLE/LSE, Box-Pierce and Ljung-Box portmanteau tests, Residual autocorrelation, Structural representation, Weak VARMA models.

Email address: `mailto:yacouba.babacarmainassara@etu.univ-lille3.fr` (Y. Boubacar Mainassara)

URL: `http://perso.univ-lille3.fr/ yboubacarmai/` (Y. Boubacar Mainassara)

1. Introduction

Consider a d-dimensional stationary process (X_t) satisfying a VARMA(p,q) representation of the form

$$A_{00}X_t - \sum_{i=1}^{p} A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{i=1}^{q} B_{0i}\epsilon_{t-i}, \quad \forall t \in \mathbb{Z} = \{0, \pm 1, \dots\}. \quad (1)$$

When $A_{00} = B_{00} = I_d$, the VARMA(p,q) representation is said to be in reduced form. Otherwise, it is said to be structural. The structural forms are mainly used in econometrics to introduce instantaneous relationships between economic variables. The representation (1) is said to satisfy a weak VARMA(p,q) representation if ϵ_t is a *weak white noise*, namely a stationary sequence of centered and uncorrelated random variables with a non singular variance Σ_0. It is customary to say that (X_t) is a strong VARMA(p,q) model if (ϵ_t) is an iid sequence of random variables with mean 0 and common variance matrix Σ_0 (*i.e. strong white noise*). A strong white noise is obviously a weak white noise, because independence entails uncorrelatedness, but the reverse is not true. Between weak and strong VARMA(p,q) representations, one can say that (1) is a semi-strong VARMA(p,q) representation if (ϵ_t) is a stationary martingale difference (*i.e. semi-strong white noise*).

The structural VARMA(p,q) representation (1) can be rewritten in a standard reduced VARMA(p,q) form if the matrices A_{00} and B_{00} are non singular. Indeed, premultiplying (1) by A_{00}^{-1} and introducing the innovation process $e_t = A_{00}^{-1}B_{00}\epsilon_t$, with non singular variance $\Sigma_{e0} = A_{00}^{-1}B_{00}\Sigma_0 B_{00}' A_{00}^{-1'}$, we obtain the reduced VARMA representation

$$X_t - \sum_{i=1}^{p} A_{00}^{-1}A_{0i}X_{t-i} = e_t - \sum_{i=1}^{q} A_{00}^{-1}B_{0i}B_{00}^{-1}A_{00}e_{t-i}. \quad (2)$$

The structural form (1) allows to handle seasonal models, instantaneous economic relationships, VARMA in the so-called echelon form representation, and many other constrained VARMA representations (see Lütkepohl, 2005, chap. 12). The reduced form (2) is more practical from a statistical viewpoint, because it gives the forecasts of each component of (X_t) according to the past values of the set of the components.

The above discussion shows that VARMA representations are not unique, that is, a given process (X_t) can be written in reduced form or in structural form by premultiplying by any non singular $(d \times d)$ matrix. Of course, in order to ensure the uniqueness of a VARMA representation, constraints are necessary for the identifiability of the $(p+q+3)d^2$ elements of the matrices involved in the VARMA equation (1). In contrast, the echelon form guarantees uniqueness of the VARMA representation (see also Lütkepohl, 2005). The echelon form is the most widely identified VARMA representation employed in the literature. The identifiability of VARMA processes has been studied in particular by Hannan (1976) who gave several procedures ensuring identifiability.

The validity of the different steps of the traditional methodology of Box and Jenkins, identification, estimation and validation, depends on the noise properties. After identification and estimation of the VARMA processes, the next important step in the VARMA modeling consists in checking if the estimated model fits satisfactorily the data. This adequacy checking step allows to validate or invalidate the choice of the orders p and q. In VARMA(p,q) models, the choice of p and q is particularly important because the number of parameters, $(p+q+2)d^2$, quickly increases with p and q, which entails statistical difficulties.

In particular, the selection of too large orders p and q has the effect of introducing terms that are not necessarily relevant in the model. Overidentification thus generally leads to a loss of precision in parameter estimation. Conversely, the selection of too small orders p and q causes loss of some information, that results in a lack of asymptotic precision for the predictions.

Thus it is important to check the validity of a VARMA(p,q) model, for given orders p and q. This paper is devoted to the problem of the validation step of VARMA representations of multivariate processes. Based on the residual empirical autocorrelations, Box and Pierce (1970) (**BP** hereafter) derived a goodness-of-fit test, the portmanteau test, for univariate strong ARMA models. Ljung and Box (1978) (**LB** hereafter) proposed a modified portmanteau test which is nowadays one of the most popular diagnostic checking tool in ARMA modeling of time series. The multivariate version of the **BP** portmanteau statistic was introduced by Chitturi (1974). We use the portmanteau tests considered by Chitturi (1974) and Hosking (1980) for checking the overall significance of the residual autocorrelations of a VARMA(p,q) model (see also Chitturi (1976), Hosking (1981a,b), Li and McLeod (1981), Ahn (1988). Hosking (1981a) gave several equivalent forms of this statistic. Arbués (2008) proposed an extended portmanteau test for VARMA models with mixing nonlinear constraints.

The works on the multivariate version of the portmanteau statistic are generally performed under the assumption that the errors ϵ_t are independent. This independence assumption is often considered too restrictive by practitioners. It precludes conditional heteroscedasticity and/or other forms of nonlinearity (see Francq and Zakoïan, 2005, for a review on weak univariate ARMA models). Relaxing this independence assumption allows to cover linear representations of general nonlinear processes and to extend the range of application of the VARMA models.

The asymptotic theory of weak ARMA model validation is mainly limited to the univariate framework (see Francq, Roy and Zakoïan, 2005, hereafter FRZ). In the multivariate analysis, a notable exception is Francq and Raïssi (2007) who study portmanteau tests for weak VAR models. We will generalize this result to VARMA models. This extension raises difficult problems. First, non trivial constraints on the parameters must be imposed for identifiability of the parameters (see Reinsel, 1997, Lütkepohl, 2005). Secondly, the implementation of standard estimation methods (for instance the Gaussian quasi-maximum likelihood estimation) is not obvious because this requires a constrained high-dimensional optimization (see also Lütkepohl, 2005). These technical difficulties certainly explain why VAR models are much more used than VARMA in applied works.

Recently Boubacar Mainassara and Francq (2010) (hereafter BMF) study the consistency and the asymptotic normality of the QMLE for a weak VARMA model. The QMLE is obtained by maximizing a function that would be the logarithm of the likelihood function if the process was Gaussian, but is not equal to it when the process (ϵ_t) is not iid Gaussian. The function that is maximized is often called quasi-likelihood. The QMLE can also be viewed as a nonlinear least squared estimator (LSE). Dufour and Pelletier (2005) and Boubacar Mainassara (in a working paper, 2010) study the choice of the orders p and q of weak VARMA models using information criteria, Chabot-Hallé and Duchesne (2008) study the asymptotic distribution of LSE and portmanteau test for semi-strong VAR. The main goal of the present paper is to complete the available results concerning the statistical analysis of weak VARMA models by considering the adequacy problem under general error terms. We proceed to study the behaviour of the goodness-of fit portmanteau tests when the ϵ_t are not independent. It is shown that the standard portmanteau tests can be quite misleading in the framework of non independent errors. Consequently, a modified version of these tests is proposed.

The paper is organized as follows. Section 2 presents the parametrization and assumptions used in the sequel. In Section 3, we recall the results on the QMLE asymptotic distribution obtained by BMF when (ϵ_t) satisfies mild mixing assumptions. Section 4 is devoted to the derivation of the joint distribution of the QMLE and the noise empirical autocovariances. In Section 5, we derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. In Section 6, it is shown how the standard Ljung-Box (or Box-Pierce) portmanteau tests must be adapted in the case of VARMA models with nonindependent innovations. Numerical experiments are presented in Section 7 and we provide a conclusion in Section 8. The proofs of the main results are collected in the appendix.

2. Parametrization and assumptions

Let $[A_{00} \dots A_{0p} B_{00} \dots B_{0q}]$ be the $d \times (p+q+2)d$ matrix of VAR and MA coefficients involved in the VARMA equation (1). The matrix Σ_0 is considered as a nuisance parameter. The parameter of interest is denoted θ_0, where θ_0 belongs to the parameter space $\Theta \subset \mathbb{R}^{k_0}$, and k_0 is the number of unknown parameters, which is typically much smaller that $(p+q+3)d^2$. The matrices $A_{00}, \dots A_{0p}, B_{00}, \dots B_{0q}$ involved in (1) and Σ_0 are specified by θ_0. More precisely, we write $A_{0i} = A_i(\theta_0)$ and $B_{0j} = B_j(\theta_0)$ for $i = 0, \dots, p$ and $j = 0, \dots, q$, and $\Sigma_0 = \Sigma(\theta_0)$. To ensure the consistence and the asymptotic normality of the QMLE, we assume that the parametrization satisfies the following smoothness conditions.

A1: The functions $\theta \mapsto A_i(\theta)$ $i = 0, \dots, p$, $\theta \mapsto B_j(\theta)$ $j = 0, \dots, q$ and $\theta \mapsto \Sigma(\theta)$ admit continuous third order derivatives for all $\theta \in \Theta$.

For simplicity, we now write A_i, B_j and Σ instead of $A_i(\theta)$, $B_j(\theta)$ and $\Sigma(\theta)$. Let $A_\theta(z) = A_0 - \sum_{i=1}^{p} A_i z^i$ and $B_\theta(z) = B_0 - \sum_{i=1}^{q} B_i z^i$. We assume that Θ corresponds to stable and invertible representations, namely

A2: for all $\theta \in \Theta$, we have $\det A_\theta(z) \det B_\theta(z) \neq 0$ for all $|z| \leq 1$.

To ensure the strong consistency of the QMLE, a compactness assumption is required.

A3: We have $\theta_0 \in \Theta$, where Θ is compact.

The structural VARMA model (1) can be written more compactly as

$$A_{\theta_0}(L)X_t = B_{\theta_0}(L)\epsilon_t \text{ where } A_{\theta_0}(L) = A_{00}-\sum_{i=1}^{p} A_{0i}L^i, B_{\theta_0}(L) = B_{00}-\sum_{i=1}^{q} B_{0i}L^i \quad (3)$$

and where L is the backward operator.

A4: The process (ϵ_t) is stationary and ergodic.

Note that **A4** is entailed by the uncorrelated innovations, but not by the iid innovations. In view of (3), $X_t = A_{\theta_0}^{-1}(L)B_{\theta_0}(L)\epsilon_t$ and $\epsilon_t = B_{\theta_0}^{-1}(L)A_{\theta_0}(L)X_t$, (ϵ_t) can be replaced by (X_t) in **A4**. In the structural VARMA model (1), the assumption **A2** does not guarantee the identifiability of the parameter. Thus, we make the following global assumption for all $\theta \in \Theta$.

A5: For all $\theta \in \Theta$ such that $\theta \neq \theta_0$, either the transfer functions $A_0^{-1}B_0B_\theta^{-1}(z)A_\theta(z) \neq A_{00}^{-1}B_{00}B_{\theta_0}^{-1}(z)A_{\theta_0}(z)$ for some $z \in \mathbb{C}$, or $A_0^{-1}B_0\Sigma B_0'A_0^{-1'} \neq A_{00}^{-1}B_{00}\Sigma_0 B_{00}'A_{00}^{-1'}$.

In the reduced VARMA representation (2), note that, the last condition in **A5** can be dropped, but may be important for structural VARMA forms. In particular, **A5** is satisfied when we impose: $A_0 = B_0 = \mathrm{I}_d$, **A2**, the common left divisors of $A_\theta(L)$ and $B_\theta(L)$ are unimodular (*i.e.* with nonzero constant determinant), and the matrix $[A_p : B_q]$ is of full rank. For the asymptotic normality of the QMLE, additional assumptions are required. It is necessary to assume that θ_0 is not on the boundary of the parameter space Θ.

A6: We have $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

We now introduce, as in Francq and Zakoïan (1998) (hereafter FZ) the strong mixing coefficients of a stationary process $Z = (Z_t)$ denoted by

$$\alpha_Z(h) = \sup_{A\in\sigma(Z_u,u\le t),B\in\sigma(X_u,u\ge t+h)} |P(A\cap B) - P(A)P(B)|\,,$$

measuring the temporal dependence of the process Z. Denoting by $\|Z\|$ the Euclidean norm of Z.

A7: We have $E\|\epsilon_t\|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$.

Note that assumption **A7** does not require independence of the noise, nor the fact that it is a martingale difference.

3. Quasi-maximum likelihood estimation

For all $\theta \in \Theta$, let $A_0 = A_0(\theta), \dots, A_p = A_p(\theta)$, $B_0 = B_0(\theta), \dots, B_q = B_q(\theta)$ and $\Sigma = \Sigma(\theta)$. Note that from **A2**, the matrices A_0 and B_0 are invertible. Thus, the structural representation (1) can be rewritten as the reduced VARMA representation (2). For the sake of simplicity, we omit the notation θ in all quantities taken at the true value θ_0. For all $\theta \in \Theta$, the assumption on the MA polynomial (from **A2**) implies that there exists a sequence of constants matrices $(C_i(\theta))$ such that $\sum_{i=1}^{\infty} \|C_i(\theta)\| < \infty$ and

$$e_t(\theta) = X_t - \sum_{i=1}^{\infty} C_i(\theta) X_{t-i}. \tag{4}$$

Given a realization $X_1, X_2, \dots, X_n$ satisfying the VARMA representation (1), the variable $e_t(\theta)$ can be approximated, for $0 < t \leq n$, by $\tilde{e}_t(\theta)$ defined recursively by

$$\tilde{e}_t(\theta) = X_t - \sum_{i=1}^{p} A_0^{-1} A_i X_{t-i} + \sum_{i=1}^{q} A_0^{-1} B_i B_0^{-1} A_0 \tilde{e}_{t-i}(\theta),$$

where the unknown initial values are set to zero: $\tilde{e}_0(\theta) = \cdots = \tilde{e}_{1-q}(\theta) = X_0 = \cdots = X_{1-p} = 0$. The Gaussian quasi-likelihood is given by

$$\tilde{\mathrm{L}}_n(\theta, \Sigma_e) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma_e}} \exp\left\{ -\frac{1}{2} \tilde{e}_t'(\theta) \Sigma_e^{-1} \tilde{e}_t(\theta) \right\}, \quad \Sigma_e = A_0^{-1} B_0 \Sigma B_0' A_0^{-1\prime}.$$

A QMLE of (θ, Σ_e) is a measurable solution $(\hat{\theta}_n, \hat{\Sigma}_e)$ of

$$(\hat{\theta}_n, \hat{\Sigma}_e) = \arg\min_{\theta, \Sigma_e} \left\{ \log(\det \Sigma_e) + \frac{1}{n} \sum_{t=1}^{n} \tilde{e}_t(\theta) \Sigma_e^{-1} \tilde{e}_t'(\theta) \right\}.$$

We use the matrix M_{θ_0} of the coefficients of the reduced form (2), where

$$M_{\theta_0} = [A_{00}^{-1} A_{01} : \cdots : A_{00}^{-1} A_{0p} : A_{00}^{-1} B_{01} B_{00}^{-1} A_{00} : \cdots : A_{00}^{-1} B_{0q} B_{00}^{-1} A_{00}].$$

Now, we need a local identifiability assumption which completes **A5** and specifies how this matrix depends on the parameter θ_0. We denote by $A \otimes B$ the Kronecker product of two matrices A and B, and by $\mathrm{vec}(A)$ the vector obtained by stacking the columns of A. Let $\dot{M}_{\theta_0}$ be the matrix $\partial \mathrm{vec}(M_\theta)/\partial \theta'$ evaluated at θ_0.

A8: The matrix $\dot{M}_{\theta_0}$ is of full rank k_0.

Under the following additional assumption, BMF showed the consistency and the asymptotic normality of the QMLE of a weak VARMA model (see Theorem 1 in BMF). One of the most popular estimation procedures is that of the least squares estimation. For the processes of the form (2), under **A1**-**A8**, it can be shown (see *e.g.* Theorem 2 in BMF), that the LSE of θ coincides with the QMLE. Then under the assumptions **A1**–**A8**, BMF showed that $\hat{\theta}_n \to \theta_0$ *a.s.* as $n \to \infty$ and $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and covariance matrix $\Sigma_{\hat{\theta}_n} := J^{-1}IJ^{-1}$, where $J = J(\theta_0, \Sigma_{e0})$ and $I = I(\theta_0, \Sigma_{e0})$, with

$$J(\theta, \Sigma_e) = \lim_{n\to\infty} \frac{-2}{n} \frac{\partial^2}{\partial\theta\partial\theta'} \log \tilde{\mathrm{L}}_n(\theta, \Sigma_e) \quad a.s.$$

and

$$I(\theta, \Sigma_e) = \lim_{n\to\infty} \mathrm{Var} \frac{2}{\sqrt{n}} \frac{\partial}{\partial\theta} \log \tilde{\mathrm{L}}_n(\theta, \Sigma_e).$$

In the standard strong VARMA case, *i.e.* when **A4** is replaced by the assumption that (ϵ_t) is an iid sequence, we have $I = 2J$, so that $\Sigma_{\hat{\theta}_n} = 2J^{-1}$.

4. Joint distribution of $\hat{\theta}_n$ and the noise empirical autocovariances

Let $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ be the quasi-maximum likelihood residuals when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$ and

$$\hat{e}_t = X_t - \sum_{i=1}^{p} A_0^{-1}(\hat{\theta}_n)A_i(\hat{\theta}_n)\hat{X}_{t-i} + \sum_{i=1}^{q} A_0^{-1}(\hat{\theta}_n)B_i(\hat{\theta}_n)B_0^{-1}(\hat{\theta}_n)A_0(\hat{\theta}_n)\hat{e}_{t-i},$$

for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$. We denote by

$$\gamma(h) = \frac{1}{n}\sum_{t=h+1}^{n} e_t e'_{t-h} \quad \text{and} \quad \hat{\Gamma}_e(h) = \frac{1}{n}\sum_{t=h+1}^{n} \hat{e}_t \hat{e}'_{t-h}$$

the white noise "empirical" autocovariances and residual autocovariances. It should be noted that $\gamma(h)$ is not a statistic (unless if $p = q = 0$) because it depends on the unobserved innovations e_t. For a fixed integer $m \geq 1$, let

$$\gamma_m = \left(\{\mathrm{vec}\gamma(1)\}', \dots, \{\mathrm{vec}\gamma(m)\}'\right)', \; \hat{\Gamma}_m = \left(\left\{\mathrm{vec}\hat{\Gamma}_e(1)\right\}', \dots, \left\{\mathrm{vec}\hat{\Gamma}_e(m)\right\}'\right)'$$

and

$$\Gamma(\ell,\ell') = \sum_{h=-\infty}^{\infty} E\left(\{e_{t-\ell}\otimes e_t\}\{e_{t-h-\ell'}\otimes e_{t-h}\}'\right), \quad \text{for} \quad (\ell,\ell')\neq(0,0).$$

For the univariate ARMA model, FRZ have showed that $\sum_{h=-\infty}^{\infty}|Ee_te_{t+\ell}e_{t+h}e_{t+h+\ell'}| < +\infty$ (see Lemma A.1), which in turn implies the existence of $\Gamma(\ell,\ell')$. We can generalize this result for the VARMA models. Then we obtain $\sum_{h=-\infty}^{\infty}\left\|E\{e_{t-\ell}\otimes e_t\}\{e_{t-h-\ell'}\otimes e_{t-h}\}'\right\| < +\infty$. The proof is similar to the univariate case.

We are now able to state the following Theorem, which is an extension of a result given in FRZ.

Theorem 4.1. *Assume that $p>0$ or $q>0$. Under Assumptions* **A1**–**A8**, *as $n\to\infty$, $\sqrt{n}(\gamma_m,\hat\theta_n-\theta_0)' \stackrel{d}{\Rightarrow} \mathcal{N}(0,\Xi)$ where*

$$\Xi = \begin{pmatrix} \Sigma_{\gamma_m} & \Sigma_{\gamma_m,\hat\theta_n} \\ \Sigma'_{\gamma_m,\hat\theta_n} & \Sigma_{\hat\theta_n} \end{pmatrix},$$

with $\Sigma_{\gamma_m} = \{\Gamma(\ell,\ell')\}_{1\le\ell,\ell'\le m}$, $\Sigma'_{\gamma_m,\hat\theta_n} = Cov(\sqrt{n}J^{-1}Y_n,\sqrt{n}\gamma_m)$ and $\Sigma_{\hat\theta_n} = \lim_{n\to\infty} Var(\sqrt{n}J^{-1}Y_n) = J^{-1}IJ^{-1}$ and Y_n is given by (15) in the proof of this Theorem. The matrices I and J are defined in Section 3.

Remark 4.1. FRZ considered the univariate case $d=1$. In their paper, they used the LSE and they obtained that

$$\Sigma_{\gamma_m} = \sum_{h=-\infty}^{+\infty}\{E(e_te_{t-\ell}e_{t-h}e_{t-\ell'-h})\}_{1\le\ell,\ell'\le m}$$

and denoted it by $\Gamma_{m,m'} = \{\Gamma(\ell,\ell')\}_{1\le\ell,\ell'\le m}$. They introduce the vectors $\lambda_i = \left(-\phi^*_{i-1},\dots,-\phi^*_{i-p},\varphi^*_{i-1},\dots,\varphi^*_{i-q}\right)' \in\mathbb{R}^{p+q}$, with the convention $\phi^*_i=\varphi^*_i=0$ when $i<0$ and where ϕ^*_h and φ^*_h denote the coefficients defined by

$$A_\theta^{-1}(z) = \sum_{h=0}^{\infty}\phi^*_h z^h, \quad B_\theta^{-1}(z) = \sum_{h=0}^{\infty}\varphi^*_h z^h, \quad |z|\le 1 \quad \text{for} \quad h\ge 0.$$

They also introduce the $(p+q)\times m$ matrices $\Lambda_m = (\lambda_1,\dots,\lambda_m)$. Using the QMLE, their result gives

$$\Sigma_{\hat\theta_n} = (\Lambda_\infty\Lambda'_\infty)^{-1}\sigma_e^{-4}\Lambda_\infty\Gamma_{\infty,\infty}\Lambda'_\infty(\Lambda_\infty\Lambda'_\infty)^{-1}$$

where σ_e^2 is the variance of the univariate process e_t. Using the fact that

$$\frac{\partial e_t(\theta_0)}{\partial \theta} = \sum_{i\geq 1} \lambda_i e_{t-i}(\theta_0),$$

we also have

$$\Sigma'_{\gamma_m,\hat{\theta}_n} = -\sigma_e^{-2} \left(\Lambda_\infty \Lambda'_\infty\right)^{-1} \Lambda_\infty \Gamma_{\infty,m},$$

which are the expressions given in Theorem 1 of FRZ.

5. Asymptotic distribution of residual empirical autocovariances and autocorrelations

Denoting the diagonal matrices by

$$S_e = \mathrm{Diag}\left(\sigma_e(1), \dots, \sigma_e(d)\right) \quad \text{and} \quad \hat{S}_e = \mathrm{Diag}\left(\hat{\sigma}_e(1), \dots, \hat{\sigma}_e(d)\right),$$

where $\sigma_e^2(i)$ is the variance of the i-th coordinate of e_t and $\hat{\sigma}_e^2(i)$ is its sample estimate (*i.e.* $\sigma_e(i) = \sqrt{Ee_{it}^2}$ and $\hat{\sigma}_e(i) = \sqrt{n^{-1}\sum_{t=1}^n \hat{e}_{it}^2}$). The theoretical and sample autocorrelations at lag ℓ are respectively defined by $R_e(\ell) = S_e^{-1}\Gamma_e(\ell)S_e^{-1}$ and $\hat{R}_e(\ell) = \hat{S}_e^{-1}\hat{\Gamma}_e(\ell)\hat{S}_e^{-1}$, with $\Gamma_e(\ell) := Ee_te'_{t-\ell} = 0$ for all $\ell \neq 0$. Consider the vector of the first m sample autocorrelations

$$\hat{\rho}_m = \left(\left\{\mathrm{vec}\hat{R}_e(1)\right\}', \dots, \left\{\mathrm{vec}\hat{R}_e(m)\right\}'\right)'.$$

The following Theorem gives the limiting distribution of the residual autocovariances and autocorrelations.

Theorem 5.1. *Under the none Assumptions as in Theorem 4.1,*

$$\sqrt{n}\hat{\Gamma}_m \Rightarrow \mathcal{N}\left(0, \Sigma_{\hat{\Gamma}_m}\right) \quad \textit{and} \quad \sqrt{n}\hat{\rho}_m \Rightarrow \mathcal{N}\left(0, \Sigma_{\hat{\rho}_m}\right) \quad \textit{where,}$$

$$\Sigma_{\hat{\Gamma}_m} = \Sigma_{\gamma_m} + \Phi_m \Sigma_{\hat{\theta}_n} \Phi'_m + \Phi_m \Sigma_{\hat{\theta}_n,\gamma_m} + \Sigma'_{\hat{\theta}_n,\gamma_m}\Phi'_m \tag{5}$$

$$\Sigma_{\hat{\rho}_m} = \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\} \Sigma_{\hat{\Gamma}_m} \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\} \tag{6}$$

and Φ_m is given by (18) in the proof of this Theorem.

Chabot-Hallé and Duchesne (2008) obtained a similar result under a difference martingale assumption on the noises.

Remark 5.1. Considered the univariate case $d=1$, as in FRZ. We obtain

$$\begin{aligned}\Phi_m &= E\left\{\begin{pmatrix} e_{t-1} \\ \vdots \\ e_{t-m}\end{pmatrix} \otimes \frac{\partial e_t'(\theta_0)}{\partial\theta}\right\} = \sum_{i\geq 1} E\left\{\begin{pmatrix} e_{t-1} \\ \vdots \\ e_{t-m}\end{pmatrix} e_{t-i}\lambda_i'\right\} \\ &= \sum_{i=1}^{m} E\left\{\begin{pmatrix} e_{t-1}e_{t-i} \\ \vdots \\ e_{t-m}e_{t-i}\end{pmatrix}\lambda_i'\right\} = \sigma_e^2\Lambda_m'.\end{aligned}$$

Using this notation and in view of Remark 4.1, Theorem 5.1 gives

$$\begin{aligned}\Sigma_{\hat{\Gamma}_m} &= \Gamma_{m,m} + \Lambda_m'\left(\Lambda_\infty\Lambda_\infty'\right)^{-1}\Lambda_\infty\Gamma_{\infty,\infty}\Lambda_\infty'\left(\Lambda_\infty\Lambda_\infty'\right)^{-1}\Lambda_m \\ &\quad -\Lambda_m'\left(\Lambda_\infty\Lambda_\infty'\right)^{-1}\Lambda_\infty\Gamma_{\infty,m} - \Gamma_{m,\infty}\Lambda_\infty'\left(\Lambda_\infty\Lambda_\infty'\right)^{-1}\Lambda_m,\end{aligned}$$

and $\Sigma_{\hat{\rho}_m} = \sigma_e^{-4}\Sigma_{\hat{\Gamma}_m}$. This last is the result given in Theorem 2 of FRZ.

Remark 5.2. In the standard strong VARMA case, *i.e.* when **A4** is replaced by the assumption that (ϵ_t) is an iid sequence, we have $\Sigma_{\gamma_m} = I_m \otimes \Sigma_{e0} \otimes \Sigma_{e0}$, $\Sigma_{\hat{\theta}_n} = 2J^{-1}$ (because $I = 2J$) and

$$\begin{aligned}\Sigma_{\hat{\theta}_n,\gamma_m}' &= -2E\left\{\begin{pmatrix} e_{t-1} \\ \vdots \\ e_{t-m}\end{pmatrix} \otimes e_t\right\}\left\{e_t'\Sigma_{e0}^{-1}\frac{\partial e_t(\theta_0)}{\partial\theta'}J^{-1}\right\} \\ &= -E\left\{\left[\begin{pmatrix} e_{t-1} \\ \vdots \\ e_{t-m}\end{pmatrix} \otimes \frac{\partial e_t(\theta_0)}{\partial\theta'}\right]\left(2J^{-1}\right)\right\} = -\Phi_m\Sigma_{\hat{\theta}_n}.\end{aligned}$$

Thus $\Sigma_{\hat{\Gamma}_m} = I_m \otimes \Sigma_{e0} \otimes \Sigma_{e0} - \Phi_m\Sigma_{\hat{\theta}_n}\Phi_m'$, that the result obtained by Hosking (1981b), Chabot-Hallé and Duchesne (2008).

6. Limiting distribution of the portmanteau statistics

Box and Pierce (1970) (**BP** hereafter) derived a goodness-of-fit test, the portmanteau test, for univariate strong ARMA models. Ljung and Box (1978) (**LB** hereafter) proposed a modified portmanteau test which is nowadays one of the most popular diagnostic checking tools in ARMA modeling of time series. The multivariate version of the **BP** portmanteau statistic was introduced by Chitturi (1974). Hosking (1981a) gave several equivalent forms of this statistic. Basic forms are

$$\begin{aligned}
Q_m &= n\sum_{h=1}^{m} \mathrm{Tr}\left(\hat{\Gamma}_e'(h)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\hat{\Gamma}_e^{-1}(0)\right) \\
&= n\sum_{h=1}^{m} \mathrm{vec}\left(\hat{\Gamma}_e(h)\right)'\left(\hat{\Gamma}_e^{-1}(0)\otimes I_d\right)\mathrm{vec}\left(\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\right) \\
&= n\sum_{h=1}^{m} \mathrm{vec}\left(\hat{\Gamma}_e(h)\right)'\left(\hat{\Gamma}_e^{-1}(0)\otimes I_d\right)\left(I_d\otimes\hat{\Gamma}_e^{-1}(0)\right)\mathrm{vec}\left(\hat{\Gamma}_e(h)\right) \\
&= n\sum_{h=1}^{m} \mathrm{vec}\left(\hat{\Gamma}_e(h)\right)'\left(\hat{\Gamma}_e^{-1}(0)\otimes\hat{\Gamma}_e^{-1}(0)\right)\mathrm{vec}\left(\hat{\Gamma}_e(h)\right) \\
&= n\hat{\Gamma}_m'\left(I_m\otimes\left\{\hat{\Gamma}_e^{-1}(0)\otimes\hat{\Gamma}_e^{-1}(0)\right\}\right)\hat{\Gamma}_m \\
&= n\hat{\rho}_m'\left(I_m\otimes\left\{\hat{\Gamma}_e(0)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(0)\right\}\otimes\left\{\hat{\Gamma}_e(0)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(0)\right\}\right)\hat{\rho}_m \\
&= n\hat{\rho}_m'\left(I_m\otimes\left\{\hat{R}_e^{-1}(0)\otimes\hat{R}_e^{-1}(0)\right\}\right)\hat{\rho}_m.
\end{aligned}$$

Where the equalities is obtained from the elementary identities $\mathrm{vec}(AB) = (I\otimes A)\,\mathrm{vec}\,B$, $(A\otimes B)(C\otimes D) = AC\otimes BD$ and $\mathrm{Tr}(ABC) = \mathrm{vec}(A')'(C'\otimes I)\,\mathrm{vec}\,B$. As for the univariate **LB** portmanteau statistic, Hosking (1980) defined the modified portmanteau statistic

$$\tilde{Q}_m = n^2\sum_{h=1}^{m}(n-h)^{-1}\mathrm{Tr}\left(\hat{\Gamma}_e'(h)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\hat{\Gamma}_e^{-1}(0)\right).$$

These portmanteau statistics are generally used to test the null hypothesis

$$H_0: \quad (X_t) \text{ satisfies a VARMA}(p,q) \text{ representation}$$

against the alternative

$$H_1: \quad (X_t) \text{ does not admit a VARMA representation or admits a VARMA}(P,Q) \text{ representation with } P>p \text{ or } Q>q.$$

These portmanteau tests are very useful tools for checking the overall significance of the residual autocorrelations. Under the assumption that the data generating process (**DGP**) follows a strong VARMA(p,q) model, the asymptotic distribution of the statistics Q_m and $\tilde{Q}_m$ is generally approximated by the $\chi^2_{d^2m-k_0}$ distribution ($d^2m > k_0$). When the innovations are gaussian, Hosking (1980) found that the finite-sample distribution of $\tilde{Q}_m$ is more nearly to $\chi^2_{d^2(m-(p+q))}$ than that of Q_m. From Theorem 5.1 we deduce the following result, in the case of weak VARMA(p,q) models, which gives the exact asymptotic distribution of the standard portmanteau statistics Q_m. We will see that the distribution may be very different from the $\chi^2_{d^2m-k_0}$ in the case of strong VARMA(p,q) models.

Theorem 6.1. *Under Assumptions in Theorem 5.1, the statistics Q_m and $\tilde{Q}_m$ converge in distribution, as $n \to \infty$, to*

$$Z_m(\xi_m) = \sum_{i=1}^{d^2m} \xi_{i,d^2m} Z_i^2$$

where $\xi_m = (\xi_{1,d^2m}, \dots, \xi_{d^2m,d^2m})'$ is the vector of the eigenvalues of the matrix

$$\Omega_m = \left(I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}\right) \Sigma_{\hat{\Gamma}_m} \left(I_m \otimes \Sigma_e^{-1/2} \otimes \Sigma_e^{-1/2}\right),$$

and $Z_1, \dots, Z_{d^2m}$ are independent $\mathcal{N}(0,1)$ variables.

Francq and Raïssi (2007) considered the sub-class of a weak vector autoregressive (VAR). They obtained a similar result and showed that the $\chi^2_{d^2(m-p)}$ approximation is no longer valid in the weak VAR(p) cases. Considering the univariate ARMA case, $d = 1$, we retrieve exactly the result given in Theorem 3 of FRZ.

It is seen in Theorem 6.1, that the asymptotic distribution of the **BP** and **LB** portmanteau tests depends on the nuisance parameters involving in Σ_e, the matrix Φ_m and the elements of the matrix Ξ. We need a consistent estimator of the above unknown matrices. The matrix Σ_e can be consistently estimated by its sample estimate $\hat{\Sigma}_e = \hat{\Gamma}_e(0)$. The matrix Φ_m can be easily estimated by its empirical counterpart

$$\hat{\Phi}_m = \frac{1}{n} \sum_{t=1}^{n} \left\{ \left(\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m}\right)' \otimes \frac{\partial e_t(\theta_0)}{\partial \theta'} \right\}_{\theta_0 = \hat{\theta}_n}.$$

In the econometric literature the nonparametric kernel estimator, also called heteroskedastic autocorrelation consistent (HAC) estimator (see Newey and West, 1987, or Andrews, 1991), is widely used to estimate covariance matrices of the form Ξ. Interpreting $(2\pi)^{-1}\Xi$ as the spectral density of the stationary process (Υ_t) evaluated at frequency 0 (see Brockwell and Davis, 1991, p. 459), an alternative method consists in using a parametric AR estimate of the spectral density of $\Upsilon_t = \left(\Upsilon'_{1,t}, \Upsilon'_{2,t}\right)'$, where $\Upsilon_{1,t} = \left(e'_{t-1}, \dots, e'_{t-m}\right)' \otimes e_t$ and $\Upsilon_{2,t} = -2J^{-1}\left(\partial e'_t(\theta_0)/\partial\theta\right)\Sigma_{e0}^{-1}e_t(\theta_0)$. This approach, which has been studied by Berk (1974) (see also den Hann and Levin, 1997), rests on the expression

$$\Xi = \mathbf{\Phi}^{-1}(1)\Sigma_u\mathbf{\Phi}'^{-1}(1)$$

when (Υ_t) satisfies an AR(∞) representation of the form

$$\mathbf{\Phi}(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty}\Phi_i\Upsilon_{t-i} = u_t, \tag{7}$$

where u_t is a weak white noise with variance matrix Σ_u. Since Υ_t is not observable, let $\hat{\Upsilon}_t$ be the vector obtained by replacing θ_0 by $\hat{\theta}_n$ in Υ_t. Let $\hat{\mathbf{\Phi}}_r(z) = \mathrm{I}_{k_0+d^2m} + \sum_{i=1}^{r}\hat{\Phi}_{r,i}z^i$, where $\hat{\Phi}_{r,1}, \dots, \hat{\Phi}_{r,r}$ denote the coefficients of the LS regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1}, \dots, \hat{\Upsilon}_{t-r}$. Let $\hat{u}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance of $\hat{u}_{r,1}, \dots, \hat{u}_{r,n}$.

We are now able to state the following Theorem, which is an extension of a result given in FRZ.

Theorem 6.2. *In addition to the assumptions of Theorem 6.1, assume that the process* (Υ_t) *admits an AR*(∞) *representation (7) in which the roots of* $\det\mathbf{\Phi}(z) = 0$ *are outside the unit disk,* $\|\Phi_i\| = o(i^{-2})$*, and* $\Sigma_u = Var(u_t)$ *is non-singular. Moreover we assume that* $E\|\epsilon_t\|^{8+4\nu} < \infty$ *and* $\sum_{k=0}^{\infty}\{\alpha_{X,\epsilon}(k)\}^{\nu/(2+\nu)} < \infty$ *for some* $\nu > 0$*, where* $\{\alpha_{X,\epsilon}(k)\}_{k\geq 0}$ *denotes the sequence of the strong mixing coefficients of the process* $(X'_t, \epsilon'_t)'$*. Then, the spectral estimator of* Ξ

$$\hat{\Xi}^{\mathrm{SP}} := \hat{\mathbf{\Phi}}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\mathbf{\Phi}}_r'^{-1}(1) \to \Xi$$

in probability when $r = r(n) \to \infty$ *and* $r^3/n \to 0$ *as* $n \to \infty$.

Let $\hat{\Omega}_m$ be the matrix obtained by replacing Ξ by $\hat{\Xi}$ and Σ_e by $\hat{\Sigma}_e$ in Ω_m. Denote by $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \dots, \hat{\xi}_{d^2m,d^2m})'$ the vector of the eigenvalues of $\hat{\Omega}_m$. At the asymptotic level α, the **LB** test (resp. the **BP** test) consists in rejecting the adequacy of the weak VARMA(p,q) model when

$$\tilde{Q}_m > S_m(1-\alpha) \quad \text{(resp.} \quad Q_m > S_m(1-\alpha))$$

where $S_m(1-\alpha)$ is such that $P\left\{Z_m(\hat{\xi}_m) > S_m(1-\alpha)\right\} = \alpha$.

7. Numerical illustrations

In this section, we present the models we simulate and we give the steps to implement the modified version of the portmanteau test. By means of Monte Carlo experiments, we investigate the finite sample properties of the test introduced in this paper.

7.1. Simulating models

To generate the strong and the weak VARMA models, we consider the following bivariate VARMA(1,1) model in echelon form

$$\begin{aligned}\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & a_1(2,2) \end{pmatrix}\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \\ &\quad - \begin{pmatrix} 0 & 0 \\ b_1(2,1) & b_1(2,2) \end{pmatrix}\begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix}, \end{aligned} \tag{8}$$

where $(a_1(2,2), b_1(2,1), b_1(2,2)) = (0.950, -0.313, 0.250)$ and $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ follows a strong or weak white noise.

7.2. Implementation of the goodness-of-fit portmanteau tests

Let $X_1, \dots, X_n$, be observations of a d-multivariate process. For testing the adequacy of a weak VARMA(p,q) model, we implement the modified version of the portmanteau test, using the following steps:

1. Compute the estimates $\hat{A}_1, \dots, \hat{A}_p, \hat{B}_1, \dots, \hat{B}_q$ by QMLE.
2. Compute the QMLE residuals $\hat{e}_t = \tilde{e}_t(\hat{\theta}_n)$ when $p > 0$ or $q > 0$, and let $\hat{e}_t = e_t = X_t$ when $p = q = 0$. When $p + q \neq 0$, we have $\hat{e}_t = 0$ for $t \leq 0$ and $t > n$ and

$$\hat{e}_t = X_t - \sum_{i=1}^{p} A_0^{-1}(\hat{\theta}_n)A_i(\hat{\theta}_n)\hat{X}_{t-i} + \sum_{i=1}^{q} A_0^{-1}(\hat{\theta}_n)B_i(\hat{\theta}_n)B_0^{-1}(\hat{\theta}_n)A_0(\hat{\theta}_n)\hat{e}_{t-i},$$

   for $t = 1, \dots, n$, with $\hat{X}_t = 0$ for $t \leq 0$ and $\hat{X}_t = X_t$ for $t \geq 1$.
3. Compute the residual autocovariances $\hat{\Gamma}_e(0) = \hat{\Sigma}_e$ and $\hat{\Gamma}_e(h)$ for $h = 1, \dots, m$ and $\hat{\Gamma}_m = \left(\left\{\hat{\Gamma}_e(1)\right\}', \dots, \left\{\hat{\Gamma}_e(m)\right\}'\right)'$.
4. Compute the matrix $\hat{J} = 2n^{-1} \sum_{t=1}^{n} \left(\partial \hat{e}_t' / \partial \theta\right) \hat{\Sigma}_{e0}^{-1} \left(\partial \hat{e}_t / \partial \theta'\right)$.

5. Compute $\hat{\Upsilon}_t = \left(\hat{\Upsilon}'_{1,t}, \hat{\Upsilon}'_{2,t}\right)'$, where $\hat{\Upsilon}_{1,t} = \left(\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m}\right)' \otimes \hat{e}_t$ and $\hat{\Upsilon}_{2,t} = -2\hat{J}^{-1}\left(\partial \hat{e}'_t/\partial\theta\right)\hat{\Sigma}_e^{-1}\hat{e}_t$.
6. Fit the VAR(r) model

$$\hat{\mathbf{\Phi}}_r(L)\hat{\Upsilon}_t := \left(I_{d^2m+k_0} + \sum_{i=1}^{r}\hat{\Phi}_{r,i}(L)\right)\hat{\Upsilon}_t = \hat{u}_{r,t}.$$

The VAR order r can be fixed in a strong VARMA case or selected by AIC/BIC information criteria in a weak VARMA case.
7. Use the spectral estimator

$$\hat{\Xi}^{\mathrm{SP}} := \hat{\mathbf{\Phi}}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\mathbf{\Phi}}_r'^{-1}(1) = \begin{pmatrix} \hat{\Sigma}_{\gamma_m} & \hat{\Sigma}_{\gamma_m,\hat{\theta}_n} \\ \hat{\Sigma}'_{\gamma_m,\hat{\theta}_n} & \hat{\Sigma}_{\hat{\theta}_n} \end{pmatrix}, \ \hat{\Sigma}_{\hat{u}_r} = \frac{1}{n}\sum_{t=1}^{n}\hat{u}_{r,t}\hat{u}'_{r,t},$$

defined in Theorem 6.2.
8. Define the estimator

$$\hat{\Phi}_m = \frac{1}{n}\sum_{t=1}^{n}\left\{\left(\hat{e}'_{t-1}, \dots, \hat{e}'_{t-m}\right)' \otimes \frac{\partial e_t(\theta_0)}{\partial\theta'}\right\}_{\theta_0=\hat{\theta}_n}.$$

9. Define the estimators

$$\begin{aligned}
\hat{\Sigma}_{\hat{\Gamma}_m} &= \hat{\Sigma}_{\gamma_m} + \hat{\Phi}_m\hat{\Sigma}_{\hat{\theta}_n}\hat{\Phi}'_m + \hat{\Phi}_m\hat{\Sigma}_{\hat{\theta}_n,\gamma_m} + \hat{\Sigma}'_{\hat{\theta}_n,\gamma_m}\hat{\Phi}'_m \\
\hat{\Sigma}_{\hat{\rho}_m} &= \left\{I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1}\right\}\hat{\Sigma}_{\hat{\Gamma}_m}\left\{I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1}\right\}.
\end{aligned}$$

10. Compute the eigenvalues $\hat{\xi}_m = (\hat{\xi}_{1,d^2m}, \dots, \hat{\xi}_{d^2m,d^2m})'$ of the matrix

$$\hat{\Omega}_m = \left(I_m \otimes \hat{\Sigma}_e^{-1/2} \otimes \hat{\Sigma}_e^{-1/2}\right)\hat{\Sigma}_{\hat{\Gamma}_m}\left(I_m \otimes \hat{\Sigma}_e^{-1/2} \otimes \hat{\Sigma}_e^{-1/2}\right).$$

11. Compute the portmanteau statistics

$$\begin{aligned}
Q_m &= n\hat{\rho}'_m\left(I_m \otimes \left\{\hat{R}_e^{-1}(0) \otimes \hat{R}_e^{-1}(0)\right\}\right)\hat{\rho}_m \quad \text{and} \\
\tilde{Q}_m &= n^2\sum_{h=1}^{m}\frac{1}{(n-h)}\mathrm{Tr}\left(\hat{\Gamma}'_e(h)\hat{\Gamma}_e^{-1}(0)\hat{\Gamma}_e(h)\hat{\Gamma}_e^{-1}(0)\right).
\end{aligned}$$

12. Evaluate the p-values

$$P\left\{Z_m(\hat{\xi}_m) > Q_m\right\} \text{ and } P\left\{Z_m(\hat{\xi}_m) > \tilde{Q}_m\right\},\ Z_m(\hat{\xi}_m) = \sum_{i=1}^{d^2m} \hat{\xi}_{i,d^2m} Z_i^2,$$

using the Imhof algorithm (1961). The **BP** test (resp. the **LB** test) rejects the adequacy of the weak VARMA(p, q) model when the first (resp. the second) p-value is less than the asymptotic level α.

7.3. Empirical size

In this paper, we only present the results of the modified and standard versions of the **LB** test. The results concerning the **BP** test are not presented here, because they are very close to those of the **LB** test. The numerical illustrations of this section are made with the softwares R (see http://cran.r-project.org/) and FORTRAN (to compute the p-values using the Imohf algorithm, 1961).

For the nominal level $\alpha = 5\%$, the empirical size over the $N = 1,000$ independent replications should vary between the significant limits 3.6% and 6.4% with probability 95%. When the relative rejection frequencies are outside the significant limits, they are displayed in bold type in Tables 1, 2, 3 and 4.

7.3.1. Strong VARMA model case

We first consider the strong VARMA case. To generate this model, we assume that in (8) the innovation process (ϵ_t) is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IID}\,\mathcal{N}(0, I_2). \tag{9}$$

We simulated $N = 1,000$ independent trajectories of size $n = 100$, $n = 500$ and $n = 2,000$ of Model (8) with the strong gaussian noise (9). For each of these N replications we estimated the coefficients $(a_1(2,2), b_1(2,1), b_1(2,2))$, using the Gaussian maximum likelihood estimation method, and we applied portmanteau tests to the residuals for different values of m, where m is the number of autocorrelations used in the portmanteau test statistic. For the standard **LB** test, the model is therefore rejected when the statistic $\tilde{Q}_m$ is greater than $\chi^2_{(4m-3)}(0.95)$. This corresponds to a nominal asymptotic level $\alpha = 5\%$ in the standard case. We know that the asymptotic level of the standard **LB** test is indeed $\alpha = 5\%$ when $(a_1(2,2), b_1(2,1), b_1(2,2)) = (0,0,0)$. Note however that, even when the noise is strong, the asymptotic level is not exactly $\alpha = 5\%$ when $(a_1(2,2), b_1(2,1), b_1(2,2)) \neq (0,0,0)$.

For the modified **LB** test, the model is rejected when the statistic $\tilde{Q}_m$ is greater than $S_m(0.95)$ *i.e.* when the p-value $(P\{Z_m(\hat{\xi}_m) > \tilde{Q}_m\})$ is less than the asymptotic level $\alpha = 0.05$. Let A and B be the (2×2)-matrices with non zero elements $a_1(2,2)$, $b_1(2,1)$ and $b_1(2,2)$. When the roots of $\det(I_2 - Az)\det(I_2 - Bz) = 0$ are near the unit disk, the asymptotic distribution of $\tilde{Q}_m$ is likely to be far from its $\chi^2_{(4m-3)}$ approximation. Table 1 displays the relative rejection frequencies of the null hypothesis H_0 that the **DGP** follows a VARMA$(1,1)$, over the $N = 1,000$ independent replications. As expected the observed relative rejection frequency of the standard **LB** test is very far from the nominal $\alpha = 5\%$ when the number m of autocorrelations used in the **LB** statistic is small. This is in accordance with the results in the literature on the standard VARMA models. In particular, Hosking (1980) showed that the statistic $\tilde{Q}_m$ has approximately the chi-squared distribution $\chi^2_{d^2(m-(p+q))}$ without any identifiability contraint. The theory that the $\chi^2_{(4m-3)}$ approximation is better for larger m is confirmed. We draw the conclusion that, even in the strong VARMA case, the modified version is preferable to the standard one, when the number m of autocorrelations used is small.

Table 1: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the strong VARMA$(1,1)$ model (8)-(9).

	$m=1$			$m=2$			$m=3$		
Length n	100	500	2,000	100	500	2,000	100	500	2,000
modified **LB**	5.5	5.6	4.0	3.7	4.4	4.8	**2.6**	4.1	**3.3**
standard **LB**	**16.2**	**16.3**	**15.5**	**8.2**	**8.0**	**7.7**	**6.7**	**6.8**	5.8
	$m=4$			$m=6$			$m=10$		
Length n	100	500	2,000	100	500	2,000	100	500	2,000
modified **LB**	**2.2**	3.9	4.4	**2.3**	3.7	**3.4**	**6.8**	4.2	**3.5**
standard **LB**	5.6	6.0	6.2	5.1	5.9	4.5	5.0	5.7	4.9

7.3.2. Weak VARMA model case

The GARCH(p,q) models constitute important examples of weak white noises in the univariate case. These models have numerous extensions to the multivariate framework (see Bauwens, Laurent and Rombouts (2006) for a review). Jeantheau (1998) has proposed a simple extension of the multivariate GARCH(p,q) with conditional constant correlation. In this model, the process (ϵ_t) verifies the following relation $\epsilon_t = H_t\eta_t$ where $\{\eta_t = (\eta_{1,t},\dots,\eta_{d,t})'\}_t$ is an iid centered process with $\mathrm{Var}\{\eta_{i,t}\} = 1$ and H_t is a diagonal matrix whose elements $h_{ii,t}$ verify

$$\begin{pmatrix} h^2_{11,t} \\ \vdots \\ h^2_{dd,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} + \sum_{i=1}^{q} A_i \begin{pmatrix} \epsilon^2_{1,t-i} \\ \vdots \\ \epsilon^2_{d,t-i} \end{pmatrix} + \sum_{j=1}^{p} B_j \begin{pmatrix} h^2_{11,t-j} \\ \vdots \\ h^2_{dd,t-j} \end{pmatrix}.$$

The elements of the matrices A_i and B_j, as well as the vector c_i, are supposed to be positive. In addition, suppose that the stationarity conditions hold. For simplicity, we consider the following bivariate ARCH(1) (*i.e.* a bivariate GARCH(p,q) model with $p = 0,\ q = 1$) model

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} h_{11,t} & 0 \\ 0 & h_{22,t} \end{pmatrix} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \tag{10}$$

where

$$\begin{pmatrix} h^2_{11,t} \\ h^2_{22,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \epsilon^2_{1,t-1} \\ \epsilon^2_{2,t-1} \end{pmatrix}.$$

We now repeat the same experiment on different weak VARMA(1,1) models. For the estimation of the coefficients, we used the quasi-maximum likelihood estimation method and we applied portmanteau tests to the residuals for different values of m. We first assume that in (8) the innovation process (ϵ_t) is an ARCH(1) model defined in equation (10) with $c_1 = 0.3$, $c_2 = 0.2$, $a_{11} = 0.45$, $a_{21} = 0.4$ and $a_{22} = 0.25$. In two other sets of experiments, we assume that in (8) the innovation process (ϵ_t) is defined by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}\eta_{2,t-1}\eta_{1,t-2} \\ \eta_{2,t}\eta_{1,t-1}\eta_{2,t-2} \end{pmatrix}, \quad \text{with} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \mathrm{IID}\,\mathcal{N}(0, I_2), \tag{11}$$

and then by

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \eta_{1,t}(|\eta_{1,t-1}|+1)^{-1} \\ \eta_{2,t}(|\eta_{2,t-1}|+1)^{-1} \end{pmatrix}, \quad \text{with} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} \sim \mathrm{IID}\,\mathcal{N}(0, I_2). \tag{12}$$

These noises are direct extensions of the weak noises defined by Romano and Thombs (1996) in the univariate case.

As expected, Tables 2 and 3 show that the standard **LB** test poorly performs to assess the adequacy of these weak VARMA$(1,1)$ models. In view of the observed relative rejection frequency, the standard **LB** test rejects very often the true VARMA$(1,1)$ and all the relative rejection frequencies are definitely outside the significant limits. By contrast, the error of first kind is well controlled by the modified version of the **LB** test. We draw the conclusion that, for these particular weak VARMA models, the modified version is clearly preferable to the standard one. In contrast, Table 4 shows that the error of first kind is well controlled by all the tests in this particular weak VARMA model, except for the standard **LB** test when $m = 1$. The modified version is also slightly preferable to the standard one.

Table 2: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA$(1,1)$ model (8)-(10).

		$m=1$			$m=2$			$m=3$	
Length n	500	2,000	10,000	500	2,000	10,000	500	2,000	10,000
modified **LB**	**6.9**	**8.5**	**7.4**	5.9	6.4	6.3	4.2	6.1	5.3
standard **LB**	**38.5**	**39.7**	**43.1**	**32.0**	**38.2**	**42.9**	**27.6**	**35.6**	**42.1**
		$m=4$			$m=6$			$m=10$	
Length n	500	2,000	10,000	500	2,000	10,000	500	2,000	10,000
modified **LB**	3.9	4.8	5.5	**3.3**	3.8	6.0	**2.7**	**3.5**	3.8
standard **LB**	**24.9**	**32.3**	**39.2**	**21.2**	**27.3**	**32.1**	**17.0**	**21.2**	**25.4**

7.4. Empirical power

In this part, we simulated $N = 1,000$ independent trajectories of size $n = 500$, $n = 1,000$ and $n = 5,000$ of a weak VARMA$(2,2)$ defined by

$$\begin{aligned}\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & 0.225 \end{pmatrix}\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0.100 \end{pmatrix}\begin{pmatrix} X_{1,t-2} \\ X_{2,t-2} \end{pmatrix} \\ &\quad + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ -0.313 & 0.250 \end{pmatrix}\begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\ &\quad - \begin{pmatrix} 0 & 0 \\ -0.140 & -0.160 \end{pmatrix}\begin{pmatrix} \epsilon_{1,t-2} \\ \epsilon_{2,t-2} \end{pmatrix}, \qquad (13)\end{aligned}$$

Table 3: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA$(1, 1)$ model (8)-(11).

	$m=1$			$m=2$			$m=3$		
Length n	500	$2,000$	$10,000$	500	$2,000$	$10,000$	500	$2,000$	$10,000$
modified **LB**	4.7	3.9	5.3	**3.4**	**2.8**	4.7	**3.1**	**2.5**	4.7
standard **LB**	**58.7**	**58.3**	**62.9**	**59.2**	**57.7**	**64.2**	**48.0**	**53.2**	**57.7**
	$m=4$			$m=6$			$m=10$		
Length n	500	$2,000$	$10,000$	500	$2,000$	$10,000$	500	$2,000$	$10,000$
modified **LB**	**2.2**	**2.2**	5.3	**1.9**	**2.0**	4.6	3.6	**3.1**	5.3
standard **LB**	**41.4**	**46.4**	**51.8**	**33.9**	**40.3**	**44.9**	**25.8**	**32.4**	**37.3**

where the innovation process (ϵ_t) is given by (11).

For each of these $N = 1,000$ replications we fitted a VARMA$(1, 1)$ model and perform standard and modified **LB** test based on $m = 1, \dots, 4$, 6 and 10 residual autocorrelations. The adequacy of the VARMA$(1, 1)$ model is rejected when the p-value is less than 5%. For this particular weak VARMA model, we have seen that the actual level of the standard version is generally much greater than the 5% nominal level (see Table 3). As in Hong (1996), we will use the empirical critical values obtained under a weak VARMA$(1, 1)$ model (8)-(11) based on $N = 1,000$ replications to compare the powers of the two tests on an equal basis. Table 5 displays the relative rejection frequencies of over the $N = 1,000$ independent replications. In this example, the standard and modified versions of the **LB** test have very similar powers. Note that, the empirical critical values strongly depend on the type of weak VARMA which is generated under the null hypothesis. Therefore, this method consisting in adjusting the critical values only works for very specific hypotheses.

8. Conclusion

In this paper we derive the asymptotic distribution of residual empirical autocovariances and autocorrelations under weak assumptions on the noise. We establish the asymptotic distribution of the **LB** (or **BP**) portmanteau test statistics for structural VARMA models with nonindependent innovations. This asymptotic distribution is quite different from the usual

Table 4: Empirical size (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA$(1,1)$ model (8)-(12).

	$m=1$			$m=2$			$m=3$		
Length n	500	2,000	10,000	500	2,000	10,000	500	2,000	10,000
modified **LB**	5.0	5.1	4.3	5.2	5.0	5.0	4.3	5.6	5.4
standard **LB**	**7.7**	**8.1**	6.4	**6.6**	5.6	6.2	5.3	6.1	5.6
	$m=4$			$m=6$			$m=10$		
Length n	500	2,000	10,000	500	2,000	10,000	500	2,000	10,000
modified **LB**	4.2	5.6	5.2	3.9	4.4	4.8	3.8	4.1	4.9
standard **LB**	4.8	6.3	5.5	4.6	4.7	4.9	4.8	4.3	4.9

chi-squared approximation (*i.e.* $\chi^2_{d^2(m-p-q)}$) used under iid assumptions on the noise. Therefore the modified versions of **LB** and **BP** are more difficult to implement because their critical values have to be computed from the data, whereas those of the standard versions are simply given in a χ^2-table.

In Monte Carlo experiments, we demonstrated that the proposed modified portmanteau test statistics have reasonable finite sample performance, at least for the models considered in our study. Under nonindependent errors, it appears that the standard test statistics are generally unreliable, overrejecting severally, while the proposed test statistics offers satisfactory levels in most cases. Even for independent errors, the modified version may be preferable to the standard one, when the number m of autocorrelations is small. Concerning the relative powers of the two versions, we also show that the modified versions of the **LB** and **BP** tests have similar powers when the critical values are adjusted. Moreover, the error of first kind is well controlled by the modified versions of the **LB** and **BP** tests. We draw the conclusion that the modified versions are preferable to the standard ones for diagnosing multivariate models under nonindependent errors.

Table 5: Empirical power (in %) of the standard and modified versions of the **LB** test in the case of the weak VARMA$(2,2)$ model (13)-(11), with critical values adjusted to obtain exactly 5% empirical sizes under the null hypothesis of the weak VARMA$(1,1)$ model (8)-(11).

	$m=1$			$m=2$			$m=3$		
Length n	500	1,000	5,000	500	1,000	5,000	500	1,000	5,000
modified **LB**	10.9	37.0	96.5	70.4	96.8	99.9	80.4	92.6	99.9
standard **LB**	9.1	14.3	99.8	54.4	87.8	100.0	53.5	87.0	100.0
	$m=4$			$m=6$			$m=10$		
Length n	500	1,000	5,000	500	1,000	5,000	500	1,000	5,000
modified **LB**	81.0	90.3	99.9	86.3	91.8	99.9	46.3	90.4	100.0
standard **LB**	50.5	86.4	100.0	50.7	87.2	100.0	45.6	87.5	100.0

9. Appendix

Proof of Theorem 4.1. Let $\tilde{\ell}_n(\theta,\Sigma_e) = -2n^{-1}\log\tilde{\mathrm{L}}_n(\theta,\Sigma_e)$. In BMF, it is shown that $\ell_n(\theta,\Sigma_e) = \tilde{\ell}_n(\theta,\Sigma_e) + o(1)$ a.s, where

$$\ell_n(\theta,\Sigma_e) := -\frac{2}{n}\log \mathrm{L}_n(\theta,\Sigma_e) = \frac{1}{n}\sum_{t=1}^{n}\left\{d\log(2\pi) + \log\det\Sigma_e + e_t'(\theta)\Sigma_e^{-1}e_t(\theta)\right\},$$

and where $(e_t(\theta))$ is given by (4). It is also shown uniformly in $\theta\in\Theta$ that

$$\frac{\partial \ell_n(\theta,\Sigma_e)}{\partial\theta} = \frac{\partial\tilde{\ell}_n(\theta,\Sigma_e)}{\partial\theta} + o(1) \quad a.s.$$

The same equality holds for the second-order derivatives of $\tilde{\ell}_n(\theta,\Sigma_e)$. In view of Theorem 1 in BMF and **A5**, we have almost surely $\hat{\theta}_n \to \theta_0 \in \overset{\circ}{\Theta}$. Thus $\partial\tilde{\ell}_n(\hat{\theta}_n,\hat{\Sigma}_e)/\partial\theta = 0$ for sufficiently large n, and a standard Taylor expansion of the derivative of $\tilde{\ell}_n$ about (θ_0,Σ_{e0}), taken at $(\hat{\theta}_n,\hat{\Sigma}_e)$, yields

$$\begin{aligned} 0 = \sqrt{n}\frac{\partial\tilde{\ell}_n(\hat{\theta}_n,\hat{\Sigma}_e)}{\partial\theta} &= \sqrt{n}\frac{\partial\tilde{\ell}_n(\theta_0,\Sigma_{e0})}{\partial\theta} + \frac{\partial^2\tilde{\ell}_n(\theta^*,\Sigma_e^*)}{\partial\theta\partial\theta'}\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) \\ &= \sqrt{n}\frac{\partial\ell_n(\theta_0,\Sigma_{e0})}{\partial\theta} + \frac{\partial^2\ell_n(\theta_0,\Sigma_{e0})}{\partial\theta\partial\theta'}\sqrt{n}\left(\hat{\theta}_n-\theta_0\right) + o_P(1), \end{aligned} \tag{14}$$

using arguments given in FZ (proof of Theorem 2), where θ^* is between θ_0 and $\hat{\theta}_n$, and Σ_e^* is between Σ_{e0} and $\hat{\Sigma}_e$, with $\hat{\Sigma}_e = n^{-1}\sum_{t=1}^n \tilde{e}_t(\hat{\theta}_n)\tilde{e}'_t(\hat{\theta}_n)$. Thus, by standard arguments, we have from (14):

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) &= -J^{-1}\sqrt{n}\frac{\partial \ell_n(\theta_0, \Sigma_{e0})}{\partial \theta} + o_P(1) \\ &= J^{-1}\sqrt{n}Y_n + o_P(1)\end{aligned}$$

where

$$Y_n = -\frac{\partial \ell_n(\theta_0, \Sigma_{e0})}{\partial \theta} = -\frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \theta}\left\{d\log(2\pi) + \log\det\Sigma_{e0} + e'_t(\theta_0)\Sigma_{e0}^{-1}e_t(\theta_0)\right\}. \quad (15)$$

Showing that the initial values are asymptotically negligible, and using well-known results on matrix derivatives (see (5) of Appendix A.13 in Lütkepohl, 2005), we have

$$Y_n = -\frac{2}{n}\sum_{t=1}^n \frac{\partial e'_t(\theta_0)}{\partial \theta}\Sigma_{e0}^{-1}e_t(\theta_0).$$

Using the elementary relation $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$ (see (4) of Appendix A.12 in Lütkepohl, 2005), we have $\operatorname{vec}\gamma(\ell) = n^{-1}\sum_{t=\ell+1}^n e_{t-\ell} \otimes e_t$. It is easily shown that for $\ell, \ell' \geq 1$,

$$\begin{aligned}\operatorname{Cov}(\sqrt{n}\operatorname{vec}\gamma(\ell), \sqrt{n}\operatorname{vec}\gamma(\ell')) &= \frac{1}{n}\sum_{t=\ell+1}^n \sum_{t'=\ell'+1}^n E\left(\{e_{t-\ell} \otimes e_t\}\{e_{t'-\ell'} \otimes e_{t'}\}'\right) \\ &\to \Gamma(\ell, \ell') \quad \text{as} \quad n \to \infty.\end{aligned}$$

Then, we have

$$\Sigma_{\gamma_m} = \{\Gamma(\ell, \ell')\}_{1 \leq \ell, \ell' \leq m}$$

By stationarity of (e_t) and (Y_t), we have

$$\begin{aligned}\operatorname{Cov}(\sqrt{n}J^{-1}Y_n, \sqrt{n}\operatorname{vec}\gamma(\ell)) &= -\frac{2}{n}\sum_{t=1}^n \sum_{t=\ell+1}^n J^{-1}\operatorname{Cov}\left(\frac{\partial e'_t(\theta_0)}{\partial \theta}\Sigma_{e0}^{-1}e_t, e_{t-\ell} \otimes e_t\right) \\ &= -\frac{2}{n}\sum_{h=-n+1}^{n-1}(n - |h|)J^{-1}\operatorname{Cov}\left(\frac{\partial e'_t(\theta_0)}{\partial \theta}\Sigma_{e0}^{-1}e_t, e_{t-h-\ell} \otimes e_{t-h}\right).\end{aligned}$$

By the dominated convergence Theorem, it follows that

$$\begin{aligned}
\mathrm{Cov}(\sqrt{n}J^{-1}Y_n, \sqrt{n}\,\mathrm{vec}\,\gamma(\ell)) &\to -\sum_{h=-\infty}^{+\infty} 2J^{-1}\mathrm{Cov}\left(\frac{\partial e_t'(\theta_0)}{\partial\theta}\Sigma_{e0}^{-1}e_t, e_{t-h-\ell}\otimes e_{t-h}\right) \\
&= -\sum_{h=-\infty}^{+\infty} 2J^{-1}E\left(\frac{\partial e_t'(\theta_0)}{\partial\theta}\Sigma_{e0}^{-1}e_t\left\{e_{t-\ell-h}\otimes e_{t-h}\right\}'\right).
\end{aligned}$$

Then we have

$$\Sigma'_{\gamma_m,\hat{\theta}_n} = -2J^{-1}\sum_{h=-\infty}^{+\infty} E\left(\frac{\partial e_t'(\theta_0)}{\partial\theta}\Sigma_{e0}^{-1}e_t\left\{\begin{pmatrix} e_{t-1-h} \\ \vdots \\ e_{t-m-h}\end{pmatrix}\otimes e_{t-h}\right\}'\right).$$

Applying the central limit Theorem (CLT) for mixing processes (see Herrndorf, 1984) we directly obtain

$$\begin{aligned}
\lim_{n\to\infty}\mathrm{Var}(\sqrt{n}J^{-1}Y_n) &= J^{-1}IJ^{-1} \\
&= \Sigma_{\hat{\theta}_n}
\end{aligned}$$

which gives the asymptotic covariance matrix of Theorem 4.1. It is clear that the existence of these matrices is ensured by the Davydov (1968) inequality. The proof is then complete. □

Proof of Theorem 5.1. Recall that

$$e_t(\theta) = X_t - \sum_{i=1}^{\infty} C_i(\theta)X_{t-i} = \mathbf{B}_\theta^{-1}(L)\mathbf{A}_\theta(L)X_t$$

where $\mathbf{A}_\theta(L) = I_d - \sum_{i=1}^{p}\mathbf{A}_iL^i$ and $\mathbf{B}_\theta(L) = I_d - \sum_{i=1}^{q}\mathbf{B}_iL^i$ with $\mathbf{A}_i = A_0^{-1}A_i$ and $\mathbf{B}_i = A_0^{-1}B_iB_0^{-1}A_0$. For $\ell = 1,\dots,p$ and $\ell' = 1,\dots,q$, let $\mathbf{A}_\ell = (a_{ij,\ell})$ and $\mathbf{B}_{\ell'} = (b_{ij,\ell'})$. We define the matrices $\mathbf{A}^*_{ij,h}$ and $\mathbf{B}^*_{ij,h}$ by

$$\mathbf{B}_\theta^{-1}(z)E_{ij} = \sum_{h=0}^{\infty}\mathbf{A}^*_{ij,h}z^h, \quad \mathbf{B}_\theta^{-1}(z)E_{ij}\mathbf{B}_\theta^{-1}(z)\mathbf{A}_\theta(z) = \sum_{h=0}^{\infty}\mathbf{B}^*_{ij,h}z^h, \quad |z|\le 1$$

for $h\ge 0$, where $E_{ij} = \partial\mathbf{A}_\ell/\partial a_{ij,\ell} = \partial\mathbf{B}_{\ell'}/\partial b_{ij,\ell'}$ is the $d\times d$ matrix with 1 at position (i,j) and 0 elsewhere. Take $\mathbf{A}^*_{ij,h} = \mathbf{B}^*_{ij,h} = 0$ when $h<0$. For any $a_{ij,\ell}$ and $b_{ij,\ell'}$, we respectively write the multivariate residual derivatives

$$\frac{\partial e_t}{\partial a_{ij,\ell}} = -\mathbf{B}_\theta^{-1}(L)E_{ij}X_{t-\ell} = -\sum_{h=0}^{\infty} \mathbf{A}^*_{ij,h}X_{t-h-\ell} \tag{16}$$

and

$$\frac{\partial e_t}{\partial b_{ij,\ell'}} = \mathbf{B}_\theta^{-1}(L)E_{ij}\mathbf{B}_\theta^{-1}(L)\mathbf{A}_\theta(L)X_{t-\ell'} = \sum_{h=0}^{\infty} \mathbf{B}^*_{ij,h}X_{t-h-\ell'}. \tag{17}$$

On the other hand, considering $\hat{\Gamma}(h)$ and $\gamma(h)$ as values of the same function at the points $\hat{\theta}_n$ and θ_0, a Taylor expansion about θ_0 gives

$$\begin{aligned}
\operatorname{vec}\hat{\Gamma}_e(h) &= \operatorname{vec}\gamma(h) + \frac{1}{n}\sum_{t=h+1}^{n}\left\{e_{t-h}(\theta)\otimes\frac{\partial e_t(\theta)}{\partial\theta'}\right. \\
&\qquad \left.+\frac{\partial e_{t-h}(\theta)}{\partial\theta'}\otimes e_t(\theta)\right\}_{\theta=\theta_n^*}(\hat{\theta}_n-\theta_0)+O_P(1/n) \\
&= \operatorname{vec}\gamma(h) + E\left(e_{t-h}(\theta_0)\otimes\frac{\partial e_t(\theta_0)}{\partial\theta'}\right)(\hat{\theta}_n-\theta_0)+O_P(1/n),
\end{aligned}$$

where θ_n^* is between $\hat{\theta}_n$ and θ_0. The last equality follows from the consistency of $\hat{\theta}_n$ and the fact that $(\partial e_{t-h}/\partial\theta')(\theta_0)$ is not correlated with e_t when $h \geq 0$. Then for $h = 1, \ldots, m$,

$$\hat{\Gamma}_m := \left(\left\{\operatorname{vec}\hat{\Gamma}_e(1)\right\}', \ldots, \left\{\operatorname{vec}\hat{\Gamma}_e(m)\right\}'\right)' = \gamma_m + \Phi_m(\hat{\theta}_n-\theta_0)+O_P(1/n),$$

where

$$\Phi_m = E\left\{\begin{pmatrix} e_{t-1} \\ \vdots \\ e_{t-m}\end{pmatrix}\otimes\frac{\partial e_t(\theta_0)}{\partial\theta'}\right\}. \tag{18}$$

In Φ_m, one can express $(\partial e_t/\partial\theta')(\theta_0)$ in terms of the multivariate derivatives (16) and (17). From Theorem 4.1, we have obtained the asymptotic joint distribution of γ_m and $\hat{\theta}_n - \theta_0$, which shows that the asymptotic distribution of $\sqrt{n}\hat{\Gamma}_m$, is normal, with mean zero and covariance matrix

$$\begin{aligned}
\lim_{n\to\infty}\operatorname{Var}(\sqrt{n}\hat{\Gamma}_m) &= \lim_{n\to\infty}\operatorname{Var}(\sqrt{n}\gamma_m) + \Phi_m\lim_{n\to\infty}\operatorname{Var}(\sqrt{n}(\hat{\theta}_n-\theta_0))\Phi_m' \\
&\quad +\Phi_m\lim_{n\to\infty}\operatorname{Cov}(\sqrt{n}(\hat{\theta}_n-\theta_0),\sqrt{n}\gamma_m) \\
&\quad +\lim_{n\to\infty}\operatorname{Cov}(\sqrt{n}\gamma_m,\sqrt{n}(\hat{\theta}_n-\theta_0))\Phi_m' \\
&= \Sigma_{\gamma_m} + \Phi_m\Sigma_{\hat{\theta}_n}\Phi_m' + \Phi_m\Sigma_{\hat{\theta}_n,\gamma_m} + \Sigma'_{\hat{\theta}_n,\gamma_m}\Phi_m'.
\end{aligned}$$

From a Taylor expansion about θ_0 of $\operatorname{vec}\hat{\Gamma}_e(0)$ we have, $\operatorname{vec}\hat{\Gamma}_e(0) = \operatorname{vec}\gamma(0) + O_P(n^{-1/2})$. Moreover, $\sqrt{n}(\operatorname{vec}\gamma(0) - E\operatorname{vec}\gamma(0)) = O_P(1)$ by the CLT for mixing processes. Thus $\sqrt{n}(\hat{S}_e \otimes \hat{S}_e - S_e \otimes S_e) = O_P(1)$ and, using (5) and the ergodic Theorem, we obtain

$$\begin{aligned}
& n\left\{\operatorname{vec}(\hat{S}_e^{-1}\hat{\Gamma}_e(h)\hat{S}_e^{-1}) - \operatorname{vec}(S_e^{-1}\hat{\Gamma}_e(h)S_e^{-1})\right\} \\
= \;& n\left\{(\hat{S}_e^{-1} \otimes \hat{S}_e^{-1})\operatorname{vec}\hat{\Gamma}_e(h) - (S_e^{-1} \otimes S_e^{-1})\operatorname{vec}\hat{\Gamma}_e(h)\right\} \\
= \;& n\left\{(\hat{S}_e \otimes \hat{S}_e)^{-1}\operatorname{vec}\hat{\Gamma}_e(h) - (S_e \otimes S_e)^{-1}\operatorname{vec}\hat{\Gamma}_e(h)\right\} \\
= \;& (\hat{S}_e \otimes \hat{S}_e)^{-1}\sqrt{n}(S_e \otimes S_e - \hat{S}_e \otimes \hat{S}_e)(S_e \otimes S_e)^{-1}\sqrt{n}\operatorname{vec}\hat{\Gamma}_e(h) \\
= \;& O_P(1).
\end{aligned}$$

In the previous equalities, we also use $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$ and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ when A and B are invertible. It follows that

$$\begin{aligned}
\hat{\rho}_m &= \left(\left\{\operatorname{vec}\hat{R}_e(1)\right\}', \dots, \left\{\operatorname{vec}\hat{R}_e(m)\right\}'\right)' \\
&= \left(\left\{(\hat{S}_e \otimes \hat{S}_e)^{-1}\operatorname{vec}\hat{\Gamma}_e(1)\right\}', \dots, \left\{(\hat{S}_e \otimes \hat{S}_e)^{-1}\operatorname{vec}\hat{\Gamma}_e(m)\right\}'\right)' \\
&= \left\{I_m \otimes (\hat{S}_e \otimes \hat{S}_e)^{-1}\right\}\hat{\Gamma}_m = \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\}\hat{\Gamma}_m + O_P(n^{-1}).
\end{aligned}$$

We now obtain (6) from (5). Hence, we have

$$\operatorname{Var}(\sqrt{n}\hat{\rho}_m) = \left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\}\Sigma_{\hat{\Gamma}_m}\left\{I_m \otimes (S_e \otimes S_e)^{-1}\right\}.$$

This completes the proof. □

Proof of Theorem 6.2. The proof is similar to that given by Francq, Roy and Zakoïan (2003) for Theorem 5.2. □

Acknowledgements

The author gratefully acknowledges the quick and careful reading of the manuscript by the two anonymous referees. Their detailed comments led to greatly improve the presentation.

References

Ahn, S. K. (1988) Distribution for residual autocovariances in multivariate autoregressive models with structured parameterization. *Biometrika* 75, 590–93.

Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Arbués, I. (2008) An extended portmanteau test for VARMA models with mixing nonlinear constraints, *Journal of Time Series Analysis* 29, 741–761.

Berk, K. N. (1974) Consistent Autoregressive Spectral Estimates. *Annals of Statistics* 2, 489–502.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006) Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–109.

Boubacar Mainassara, Y. (2010) Selection of weak VARMA models by modified Akaike's information criteria. *Working Papers,* http://mpra.ub.uni-muenchen.de/24981/.

Boubacar Mainassara, Y. and Francq, C. (2010) Estimating structural VARMA models with uncorrelated but non-independent error terms. *Journal of Multivariate Analysis,* to appear.

Box, G. E. P. and Pierce, D. A. (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–26.

Brockwell, P. J. and Davis, R. A. (1991) *Time series: theory and methods.* Springer Verlag, New York.

Chabot-Hallé, D. and Duchesne, P. (2008) Diagnostic checking of multivariate nonlinear time series models with martingale difference errors, *Statistics and Probability Letters* 78, 997–1005.

Chitturi, R. V. (1974) Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association* 69, 928–934.

Chitturi, R. V. (1976) Distribution of multivariate white noise autocorrelations. *Journal of the American Statistical Association* 71, 223–226.

Davydov, Y. A. (1968) Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability and Applications* 13, 691–696.

den Hann, W. J. and Levin, A. (1997) A Practitioner's Guide to Robust Covariance Matrix Estimation. *In Handbook of Statistics* 15, Rao, C.R. and G.S. Maddala (eds), 291–341.

Dufour, J-M. and Pelletier, D. (2005) Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application. *Technical report, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada.*

Francq, C. and Raïssi, H. (2007) Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors, *Journal of Time Series Analysis* 28, 454–470.

Francq, C., Roy, R. and Zakoïan, J-M. (2003) Goodness-of-fit tests for ARMA models with uncorrelated errors, *Discussion Paper CRM-2925, Centre de recherches mathématiques, Université de Montréal.* http://www.crm.umontreal.ca/pub/Rapports/2900-2999/2925.pdf

Francq, C., Roy, R. and Zakoïan, J-M. (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.

Francq, C. and Zakoïan, J-M. (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.

Francq, C. and Zakoïan, J-M. (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems,* Chap. 12 (eds P. Duchesne and B. Rémillard). New York: Springer Verlag, 241–265.

Hannan, E. J. (1976) The identification and parametrization of ARMAX and state space forms, *Econometrica* 44, 713–723.

Herrndorf, N. (1984) A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability* 12, 141–153.

Hong, Y. (1996) Consistent testing for serial correlation of unknown form, *Econometrica* 64, 837–864.

Hosking, J. R. M. (1980) The multivariate portmanteau statistic, *Journal of the American Statistical Association* 75, 602–608.

Hosking, J. R. M. (1981a) Equivalent forms of the multivariate portmanteau statistic, *Journal of the Royal Statistical Society* B 43, 261–262.

Hosking, J. R. M. (1981b) Lagrange-tests of multivariate time series models, *Journal of the Royal Statistical Society* B 43, 219–230.

Imhof, J. P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* 14, 70–86.

Li, W. K. and McLeod, A. I. (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models, *Journal of the Royal Statistical Society* B 43, 231–239.

Ljung, G. M. and Box, G. E. P. (1978) On measure of lack of fit in time series models. *Biometrika* 65, 297–303.

Lütkepohl, H. (2005) *New introduction to multiple time series analysis.* Springer Verlag, Berlin.

Newey, W. K. and West, K. D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Reinsel, G. C. (1997) *Elements of multivariate time series Analysis.* Second edition. Springer Verlag, New York.

Romano, J. L. and Thombs, L. A. (1996) Inference for autocorrelations under weak assumptions, *Journal of the American Statistical Association* 91, 590–600.